PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Evaluating the development, woman-centricity and psychometric properties of maternity patient-reported outcome measures (PROMs) and patient-reported experience measures (PREMs): A systematic review protocol
AUTHORS	Bull, Claudia; Teede, Helena; Carrandi, Lane; Rigney, Azure; Cusack, Sally; Callander, Emily

VERSION 1 – REVIEW

REVIEWER	Greene, Richard University College Cork, National Perinatal Epidemiology Centre
REVIEW RETURNED	11-Nov-2021

GENERAL COMMENTS	This proposed protocol is a valuable idea and would provide a useful assessment of available tools with a subsequent database. There is a sense in reading this protocol that there are strong views about what makes a good tool that may be a conflict of interest to achieving a robust scientific output from this work. These tools need to be considered with respect to woman centeredness - this appers to be equated to being fulfilled only by woman involvement in the design process - that may lead to many valuable tools being discarded. An assessment of this involvement is appropriate and necessary and the authors likely need to identify how best to assess it, measure it and highlight how it alters the value of the instrument. Many of the PROMs and PREMs may in fact be very woman centred without having involved woman service users in the design and development - that should not dismiss the tool but the issue should be highlighted. Thee is real value in looking at PROMs and PREMs as often both are needed to provide a full assessment of the woman's health and wellbeing and experience of the care while achieving that status. There are a number of issues that need to be reviewed and considered in this manuscript. These are detailed below The title suggests a systematic review of PROMs and PREMs supporting the development of a woman centered instrument database; yet the abstract introduction appears to have a different endpoint - This information will be used to develop a maternity PROMs and PREMs database to support service and system performance measurement, and value-based maternity care initiatives. They are not mutually exclusive but have a different focus and PROM's assist care management while PREM's relate to experience; value based maternity care is again a different focus - the authors need to be clear. The idea is a valuable idea to develop such a database – it will be more valuable if each tool type PROM, PREM, etc is clearly delineated.

	The opening paragraph on page 6/27 is a statement around disciplines, it is political, lazy and while it gets to a valuable issue of women's values and preferences – the effect is lost by cantering it on interdisciplinary…… Woman centred care is not the preserve of any one discipline. The focus should be on women and WCC – let their inputs tell us about the disciplines. This introduction is not helpful to the valuable idea behind the manuscript. Page6/27 (4) - Line 57-59 – mentions PROMs and PREMS but describe the use of PREMs – corrected in lines 3-24 page 7/27 There is a a lack of clarity in the introduction around PROMS/PREMs and value – this needs to be very clear PROMs are self-report questionnaires, completed by patients, and seek to measure their perceptions of their health status and health-related quality of life. Although variable in application to a population or to a specific condition, the content tends to focus on one or more of the following: physical functioning, symptoms, social wellbeing, psychological wellbeing, cognitive function and role activities. Patient outcomes measure changes to patient perceived health status after care interventions. A PREM is a measure of a patient's perception of their personal experience of the healthcare they have received. PREM instruments should focus on the aspects of the care that matter to the patient. PREM results can be used to improve services and provide a patient view on these improvements that moves away from the technological or economic model that is often employed in service design. Patient experience measures provide an insight into the quality of the care as experienced by the patient. A reasonable recommendation is that both PROMs and PREMs are used in order to provide a new perspective to both healthcare managers and clinicians which emphasises that services should put patients at the centre and not the priorities of clinicians or managers. I am unclear from the text if the authors are clear on the differences? Page 7/27 (5) Line 31 – 'woman-centric instrument development' this needs to be clarified and does not fit with the reference provided. Does it mean women as service users have been involved in developing the content? Methods and Analysis: Abstract and page 8/27 'The COSMIN risk of bias checklist will be used to evaluate the quality of studies reporting on the development, content validation, and/or psychometric evaluation of PROMs and PREMs. COSMIN criteria for good content validity will be used to assess the woman-centricity of PROM and PREM development and content validation studies' – Similar discussions in the main body of the paper – Cosmin was developed to assess PROMs – it may be used for PREM's but in a limited way – I would suggest a review of the following: 'COSMIN checklists have been designed and validated for use in evaluating the rigour of psychometric studies of healthcare instruments. The COSMIN checklists only enable a critique of the validity and reliability aspects of utility; additional questions to rate the cost efficiency, acceptability and educational impact of instruments maybe needed for PREM's' - see Beattie et al. Systematic Reviews (2015).

	However, the COSMIN checklist for content validity does not cover all aspects of user acceptability, e.g. cognitive testing Page 8/27 (6) – from line28 Eligibility Criteria: Why is the cut-off for publications at 2010? Yet older instruments will be considered – perhaps the authors would consider if they will review a longer period or not. As they are looking at a database development – perhaps a more holistic view of the available tools should be considered? Line 47 – only PREM's mentioned and they are not the tool for 'psychometric evaluation' Page 9/27 (7) Studies excluded if: Line 12/13 – 'designed for a non-maternity care context' - this may exclude valuable tools usable in women in maternity care i.e. Depression Scales Line 23-28 – exclusions again rule out valuable tools which may have a place in the database? Some of the exclusions are for women undergoing pregnancy related events? Line 36-37 – 'Described PROMs or PREMs designed for care providers, children, or proxies (e.g., partner or carer completes the PROM or PREM on behalf of women' these are not PROMs or PREMS - so you can state 'tools that are not Patient reported. Line 40-43 – HRQoL – these should be included as per the inclusion criteria? HRQoL tools are PROM's and valuable – the discussion to exclude them is not valid. QoL is an assessment of a health state and it is widely used in research and not confined to cost utility it is much broader. Page 10/27 (8) Line 39 – using the search terms in Cosmin for PROMs may lead to a suboptimal identification of PREM's – while (9/27) discussion took place appropriately of not assessing satisfaction studies – there may be value in searching for these tools because as stated they get mixed up with PREM's? Page 11/27 (9) Line 56-60 – one reviewer will assess if other researchers evaluated risks of bias – this step is dubious. The researchers should do a full assessment as outlined earlier – other researchers may have been wrong or viewing the instruments from a different viewpoint? The research team should do their own risk bias for every paper? Page 13/27 (11) Line 3-5 – while appreciateing the researchers wish to view the involvement of women (service users assumed) – labelling those not demonstrating such as 'inadequate' is inappropriate – that is not content validity. It probably warrants an additional score in some way as PREM's cannot be fully assessed with just the Cosmin toolkit? Lines 28-31 – same issue as above and secondly one reviewer undertaking the content validation assessment – Is this appropriate considering Bias, etc? Here the authors are talking about 'woman-centeredness' – is this the same as or confused with involvement of women – they are completely different constructs but it is difficult to decide if the authors mean the same or different things? Page 14/27 (12) Line 42-43 - One reviewer undertaking the assessment – concerns for bias

	Page 15/27 (13) Line 55 – Interpretability and feaseability – there needs to be a clear understanding how this is to be measured and shown – is this not covered within content validity, psychometric properties. Again with PREM's - the COSMIN checklists only enable a critique of the validity and reliability aspects of utility; additional questions to rate the cost efficiency, acceptability and educational impact of instruments maybe needed for PREM's' - see Beattie et al. Systematic Reviews (2015). Page 18/27 (16) Conclusion – The focus on the extent of wome's involvement in the development and content validation is really important but the authors appear to have a very significant focus on this one issue. That focus is a bias before commencing the work. The degree of woman involvement should be clearly assess, where it puts ann instrument at risk of not performing well that should be called out – it should not necessarily preclude valuable tools but within the database – the deficit should be highlighted and that then allows potetntial future users to see the deficit and perhaps alter the tool to ensure enhanced woman centeredness. The grading of the quality of the evidence has not been considered in this manuscript.

REVIEWER	Dawson, Pauline University of Otago, Women's & Children's Health
REVIEW RETURNED	22-Nov-2021

GENERAL COMMENTS	Thank you for the opportunity to review this study protocol. It is encouraging to see this topic being examined as Patient-Reported Outcome Measures (PROMs) and Patient-Reported Experience Measures (PREMs) are vital to service quality improvement in the maternity field. The need for this is described well in the introduction and development of an open-access database an admirable aim. In more specific points, the eligibility criteria states only studies published in English will be included. This should be included as a limitation. While this decision is likely resource-dependent, it has been noted that non-English language as an exclusion criterion in reviews can introduce a risk of bias (1,2). Several mentions are made of different circumstances, subgroups, and heterogeneous values and preferences. I would like to see a clearer discussion in the methods about how the data about PREMS and PROMS will be evaluated across a diverse range of birthing populations. Looking at international clinical maternity outcomes there is clear evidence of inequity, particularly in Black, Indigenous, and People of Colour (BIPOC) populations. It would be important to see how this is revealed in PROMS. Also, while the authors specifically discuss the difference between PREMs and satisfaction surveys on page 8, inequity has also been shown in satisfaction measures. Discussions of PREMs in the context of a range of populations is also vital. As the authors discuss in lines 45-54 on page 4, it is likely PREMs and PROMs will not be standard for all, not just in different settings but acknowledging the birthing population is not homogenous. The COSMIN cross-cultural validity measurement (Table 1) may go some way to address this but additional clarity would be helpful to this reader.

	A minor comment is that there are the changes in tense in the document e.g. Page 5 Line 58 “this study will be “compared to page 6 line 32 “studies were included” The public and patient involvement in the project is excellent to see. I assume this would include diverse representation and cross-cultural consumer input. A comment around consumer inclusivity would be useful. 1. Neimann Rasmussen, L., Montgomery, P. The prevalence of and factors associated with inclusion of non-English language studies in Campbell systematic reviews: a survey and meta-epidemiological study. Syst Rev 7, 129 (2018). https://doi.org/10.1186/s13643-018-0786-6 2. Morrison, A., Polisena, J., Husereau, D., Moulton, K., Clark, M., Fiander, M., . . . Rabb, D. (2012). THE EFFECT OF ENGLISH-LANGUAGE RESTRICTION ON SYSTEMATIC REVIEW-BASED META-ANALYSES: A SYSTEMATIC REVIEW OF EMPIRICAL STUDIES. International Journal of Technology Assessment in Health Care, 28(2), 138-144. doi:10.1017/S0266462312000086

REVIEWER	Sultan, P Stanford University
REVIEW RETURNED	22-Nov-2021

GENERAL COMMENTS	I think this is a nicely written and timely piece as there is much interest in the use of peri-partum patient reported measures. The review of peripartum PREMs is novel and interesting. My one reservation is that the authors do not describe how their work differs (or even acknowledge the publication of) several recent COSMIN publications in the postpartum space: PMID: 32442292 PMID: 34042993 PMID: 34013345 PMID: 34016441 PMID: 31429919 The authors should be strongly encouraged to explicitly state how their work is different to that published already and what gaps exist in the current literature that will be answered by their review.

VERSION 1 – AUTHOR RESPONSE

Reviewer 1	
R4: This proposed protocol is a valuable idea and would provide a useful assessment of available tools with a subsequent database. There is a sense in reading this protocol that there are strong views about what makes a good tool that may be a conflict of interest to achieving a robust scientific output from this work. These tools need to be considered with respect to woman centeredness - this appears to be	A4: Thank you for taking the time to provide feedback on our manuscript. We will not be excluding instruments that have failed to involve women in their conceptualisation or content validation. Instead, we note the critical importance of involving women in determining the content

equated to being fulfilled only by woman involvement in the design process - that may lead to many valuable tools being discarded. An assessment of this involvement is appropriate and necessary and the authors likely need to identify how best to assess it, measure it and highlight how it alters the value of the instrument. Many of the PROMs and PREMs may in fact be very woman centred without having involved woman service users in the design and development - that should not dismiss the tool but the issue should be highlighted. There is real value in looking at PROMs and PREMs as often both are needed to provide a full assessment of the woman's health and wellbeing and experience of the care while achieving that status. There are a number of issues that need to be reviewed and considered in this manuscript. These are detailed below	of instruments to support woman-centred, value-based measurement. We detail this further below.
R5: The title suggests a systematic review of PROMs and PREMs supporting the development of a woman centered instrument database; yet the abstract introduction appears to have a different endpoint - This information will be used to develop a maternity PROMs and PREMs database to support service and system performance measurement, and value-based maternity care initiatives. They are not mutually exclusive but have a different focus and PROM's assist care management while PREM's relate to experience; value based maternity care is again a different focus - the authors need to be clear. The idea is a valuable idea to develop such a database – it will be more valuable if each tool type PROM, PREM, etc is clearly delineated.	A5: We have now revised the title to: “Evaluating the development, woman-centricity and psychometric properties of maternity patient-reported outcome measures (PROMs) and patient-reported experience measures (PREMs): A systematic review protocol”
R6: The opening paragraph on page 6/27 is a statement around disciplines, it is political, lazy and while it gets to a valuable issue of women’s values and preferences – the effect is lost by cantering it on interdisciplinary…… Woman centred care is not the preserve of any one discipline. The focus should be on women and WCC – let their inputs tell us about the disciplines. This introduction is not helpful to the valuable idea behind the manuscript.	A6: In line with the feedback received by reviewers 2 and 3, we have kept paragraph one in its original form. The intention of this paragraph is to highlight that woman-centred care should be interdisciplinary, and important for all women regardless of who is leading their care. We have revised other sections of the Introduction; please see below comments for details.
R7: Page6/27 (4) - Line 57-59 – mentions PROMs and PREMS but describe the use of PREMs – corrected in lines 3-24	A7: Both PROMs and PREMs contribute to evaluating health services and systems performance. We have now provided reference to the OECD to support this (please refer to pg. 5, reference 15).
R8: page 7/27 There is a a lack of clarity in the introduction around PROMS/PREMs and value – this needs to be very clear PROMs are self-report questionnaires, completed by patients, and seek to measure their perceptions of their health status and health-related quality of life. Although variable in	A8: We had now revised the introduction to more clearly articulate how PROMs and PREMs contribute to value, and to make clearer the distinction between PROMs and PREMs (please refer to pg. 4-5).

application to a population or to a specific condition, the content tends to focus on one or more of the following: physical functioning, symptoms, social wellbeing, psychological wellbeing, cognitive function and role activities. Patient outcomes measure changes to patient perceived health status after care interventions. A PREM is a measure of a patient's perception of their personal experience of the healthcare they have received. PREM instruments should focus on the aspects of the care that matter to the patient. PREM results can be used to improve services and provide a patient view on these improvements that moves away from the technological or economic model that is often employed in service design. Patient experience measures provide an insight into the quality of the care as experienced by the patient. A reasonable recommendation is that both PROMs and PREMs are used in order to provide a new perspective to both healthcare managers and clinicians which emphasises that services should put patients at the centre and not the priorities of clinicians or managers. I am unclear from the text if the authors are clear on the differences?	
R9: Page 7/27 (5) Line 31 – 'woman-centric instrument development' this needs to be clarified and does not fit with the reference provided. Does it mean women as service users have been involved in developing the content?	A9: The reference provided (Terwee et al, 2018) appropriately articulates the importance of involving patients (i.e., women, in the context of maternity care) in defining what is relevant, comprehensive and comprehendible instrument content. We have now clarified what woman-centric instrument development and content validation refers to: "… woman-centric instrument development and content validation – that is, the involvement of women in defining what is relevant, comprehensive and comprehendible instrument content – is crucial to supporting meaningful, value-based measurement in maternity care.(16)" (pg. 5)
R10: Abstract and page 8/27 'The COSMIN risk of bias checklist will be used to evaluate the quality of studies reporting on the development, content validation, and/or psychometric evaluation of PROMs and PREMs. COSMIN criteria for good content validity will be used to assess the woman-centricity of PROM and PREM development and content validation studies' –	A10: We have already noted that "a potential limitation of this review is using COSMIN guidance (developed for PROMs) to evaluate the development, content validation and psychometric evaluation of PREMs." (pg. 3) Given the current lack of PREM-specific guidance and similarities in how PROMs and PREMs

Similar discussions in the main body of the paper – Cosmin was developed to assess PROMs – it may be used for PREM's but in a limited way – I would suggest a review of the following: 'COSMIN checklists have been designed and validated for use in evaluating the rigour of psychometric studies of healthcare instruments. The COSMIN checklists only enable a critique of the validity and reliability aspects of utility; additional questions to rate the cost efficiency, acceptability and educational impact of instruments maybe needed for PREM's' - see Beattie et al. Systematic Reviews (2015). However, the COSMIN checklist for content validity does not cover all aspects of user acceptability, e.g. cognitive testing Page 8/27 (6) – from line28	have been designed and evaluated over recent decades, COSMIN in the best available evidence for assessing PREM psychometric properties. In our assessment of interpretability and feasibility, we address Beattie's issues of cost efficiency and acceptability. As currently noted on pg. 14, we will be extracting the following feasibility data from studies: "(i) available modes of administration; (ii) length of the instrument; (iii) estimated completion time; (iv) level of readability; (v) ease of response calculation; (vi) copyright; (vii) cost of using an instrument; (viii) equipment required for instrument administration; (ix) availability of instrument for application in different settings and languages; and (x) approvals required before instrument use." This addresses the costs of implementing PROMsnd PREMs in practice. Additionally, we will be extracting data such as the number of items and proportion of missing data, alongside content validity in order to demonstrate aspects of acceptability. Educational impact, as described by Beattie refers to "How easy is it for an organisation, or individual within it, to drill down and make use of the data?". This is a different aspect of PROMs and PREMs that relates to using the data to inform service improvement. Thus, it is outside of the scope of this review, but an interesting and noteworthy consideration for further research.
R11: Eligibility Criteria: Why is the cut-off for publications at 2010? Yet older instruments will be considered – perhaps the authors would consider if they will review a longer period or not. As they are looking at a database development – perhaps a more holistic view of the available tools should be considered? Line 47 – only PREM's mentioned and they are not the tool for 'psychometric evaluation'	A11: We have revised the eligibility criteria for our review, please see pg. 7-8. We will only be including instruments published from 2010 onwards as these represent contemporary instruments. We will include development and psychometric evidence for the identified instruments that pre-dates 2010 by hand-searching the reference lists of included articles. We have now acknowledged this as a limitation on pg. 3.
R12: Page 9/27 (7) Studies excluded if: Line 12/13 – 'designed for a non-maternity care context' - this may exclude valuable tools usable in women in maternity care i.e. Depression Scales Line 23-28 – exclusions again rule out valuable tools which may	A12: We have revised the eligibility criteria for our review, please see pg. 7-8. Our aim is to identify instruments that capture outcomes and experiences relevant to all women across the pregnancy, childbirth

have a place in the database? Some of the exclusions are for women undergoing pregnancy related events?	and postpartum continuum. Specifically, we want to identify those developed in the maternity context because it is currently unclear (i) whether these exist, and (ii) are 'good' instruments. There is already significant evidence to support generic instruments like PROMIS and SF-36. However, given the heterogeneity of maternity populations, it may be that generic measures do not suitably capture the nuances of maternity outcomes and experiences. Thus, this needs to be further investigated.
R13: Line 36-37 – 'Described PROMs or PREMs designed for care providers, children, or proxies (e.g., partner or carer completes the PROM or PREM on behalf of women' these are not PROMs or PREMS - so you can state 'tools that are not Patient reported.	A13: This has been revised to: "Described proxy-reported PROMs/ PREMs (i.e., not self-reported by women)" pg.7
R14: Line 40-43 – HRQoL – these should be included as per the inclusion criteria? HRQoL tools are PROM's and valuable – the discussion to exclude them is not valid. QoL is an assessment of a health state and it is widely used in research and not confined to cost utility it is much broader.	A14: We disagree with the reviewer and have provided several references in-text that similarly acknowledge the difference between HRQoL measures and PROMs. To make it clearer to the reader exactly the types of instruments we are referring to, we have provided examples of quality of life instruments and have changed HRQoL to quality of life/ utility measures. (pg. 9)
R15: Page 10/27 (8) Line 39 – using the search terms in Cosmin for PROMs may lead to a suboptimal identification of PREM's – while (9/27) discussion took place appropriately of not assessing satisfaction studies – there may be value in searching for these tools because as stated they get mixed up with PREM's?	A15: As illustrated in the supplementary file 1, we have run different search strategies for PROMs and PREMs using different terms for each type of instrument, but employing the same COSMIN search terms relevant to measurement properties.
R16: Page 11/27 (9) Line 56-60 – one reviewer will assess if other researchers evaluated risks of bias – this step is dubious. The researchers should do a full assessment as outlined earlier – other researchers may have been wrong or viewing the instruments from a different viewpoint? The research team should do their own risk bias for every paper?	A16: COSMIN recommends that if others have already rated the quality of PROM development, then it should not be done again: "CHECK EXISTING RATINGS of the quality of the PROM development Step 1 (evaluating the quality of the PROM development) needs to be done only once per PROM. Ratings of the quality of PROM developments are collected and published on the COSMIN website. We recommend to check the COSMIN website first to see if the quality of the PROM development has already been rated (e.g. in another systematic review). If a rating of the PROM development already exists, we

	recommend reviewers to consider using this rating instead of rating the quality of the PROM development again." (COSMIN methodology for assessing the content validity of PROMs – User manual, pg. 16) However, if additional evidence has been published for a PROM since previous researchers' ratings, then we will also take that into consideration (please see revision on manuscript pg. 11).
R17: Page 13/27 (11) Line 3-5 – while appreciating the researchers wish to view the involvement of women (service users assumed) – labelling those not demonstrating such as 'inadequate' is inappropriate – that is not content validity. It probably warrants an additional score in some way as PREM's cannot be fully assessed with just the Cosmin toolkit?	A17: The COSMIN methodology for assessing content validity has 5 standards for evaluating the quality of studies on the content validity of instruments: 1. Asking patients about the relevance of items 2. Asking patients about the comprehensiveness of the instrument 3. Asking patients about the comprehensibility of the instrument 4. Asking professionals about the relevance of items 5. Asking professionals about the comprehensiveness of the instrument If patients (i.e., women) were not asked about the relevance, comprehensiveness or comprehensibility of the instrument, this will be automatically scored 'inadequate' according to the COSMIN guidance. As such, this an appropriate label for these studies.
R18: Lines 28-31 – same issue as above and secondly one reviewer undertaking the content validation assessment – Is this appropriate considering Bias, etc? Here the authors are talking about 'woman-centeredness' – is this the same as or confused with involvement of women – they are completely different constructs but it is difficult to decide if the authors mean the same or different things?	A18: Please see our earlier comments about clarifying woman-centricity and the content validity assessment. We will now have 2 reviewers undertaking content validity assessment: "Two reviewers will undertake the content validation assessment, and inter-rater reliability scores will be reported." (pg. 12)
R19: Page 14/27 (12) Line 42-43 - One reviewer undertaking the assessment – concerns for bias	A19: We will now have 2 reviewers undertaking psychometric properties assessment: "Two reviewers will undertake the good psychometric properties assessment." (pg. 14)
R20: Page 15/27 (13) Line 55 – Interpretability and feasibility – there needs to be a clear understanding how this is to be measured and shown – is this not covered within content validity, psychometric properties. Again with PREM's - the	A20: As is stated in the protocol, we will describe the interpretability and feasibility of instrument implementation based on the information provided within the included studies. We will not

COSMIN checklists only enable a critique of the validity and reliability aspects of utility; additional questions to rate the cost efficiency, acceptability and educational impact of instruments maybe needed for PREM's' - see Beattie et al. Systematic Reviews (2015).	be measuring interpretability or feasibility as that is beyond the scope of this work. We have clearly articulated how interpretability and feasibility are defined and what data we will be extracting to demonstrate those aspects of the included maternity PROMs and PREMs (please refer to pg. 15-16). Please see our earlier comments with regards to Beattie et al.
R21: Page 18/27 (16) Conclusion – The focus on the extent of women's involvement in the development and content validation is really important but the authors appear to have a very significant focus on this one issue. That focus is a bias before commencing the work. The degree of woman involvement should be clearly assess, where it puts and instrument at risk of not performing well that should be called out – it should not necessarily preclude valuable tools but within the database – the deficit should be highlighted and that then allows potential future users to see the deficit and perhaps alter the tool to ensure enhanced woman centeredness.	A21: Please refer to our earlier comments.
R22: The grading of the quality of the evidence has not been considered in this manuscript.	A22: We have now included a paragraph on summarising and grading the quality of evidence: "By summarising and grading the evidence available for an individual instrument, we can provide an overall conclusion as to the quality of that instrument. Thus, this will involve combining the results of each instruments' risk of bias, content validity, and psychometric property assessments into a single metric of 'high', 'moderate', 'low', or 'very low' evidence using the Grading of Recommendations Assessment, Development, and Evaluation (GRAE) approach.(29) If the results across multiple studies pertaining to a single instrument are consistent, then results will be quantitatively pooled and a GRADE score will be reported. If results are inconsistent, they will not be pooled, no GRADE score will be reported, and areas of inconsistency will be discussed (e.g., if an instrument demonstrates differing levels of quality depending on the country in which it is used)." (pg. 15)
Reviewer 2	
R23: Thank you for the opportunity to review this study protocol. It is encouraging to see this topic being examined as Patient-Reported Outcome Measures (PROMs) and	A23: Thank you for taking the time to review our manuscript and for recognising its importance.

Patient-Reported Experience Measures (PREMs) are vital to service quality improvement in the maternity field. The need for this is described well in the introduction and development of an open-access database an admirable aim.	
R24: In more specific points, the eligibility criteria states only studies published in English will be included. This should be included as a limitation. While this decision is likely resource-dependent, it has been noted that non-English language as an exclusion criterion in reviews can introduce a risk of bias (1,2). 1. Neimann Rasmussen, L., Montgomery, P. The prevalence of and factors associated with inclusion of non-English language studies in Campbell systematic reviews: a survey and meta-epidemiological study. Syst Rev 7, 129 (2018). https://doi.org/10.1186/s13643-018-0786-6 2. Morrison, A., Polisena, J., Husereau, D., Moulton, K., Clark, M., Fiander, M., . . . Rabb, D. (2012). THE EFFECT OF ENGLISH-LANGUAGE RESTRICTION ON SYSTEMATIC REVIEW-BASED META-ANALYSES: A SYSTEMATIC REVIEW OF EMPIRICAL STUDIES. International Journal of Technology Assessment in Health Care, 28(2), 138-144. doi:10.1017/S0266462312000086	A24: We have now added this as a limitation on pg. 3.
R25: Several mentions are made of different circumstances, subgroups, and heterogeneous values and preferences. I would like to see a clearer discussion in the methods about how the data about PREMS and PROMS will be evaluated across a diverse range of birthing populations.	A25: We have included the following with relation to discussing inconsistent results related to the overall quality of included instruments: “If the results across multiple studies pertaining to a single instrument are consistent, then results will be quantitatively pooled and a GRADE score will be reported. If results are inconsistent, they will not be pooled, no GRADE score will be reported, and areas of inconsistency will be discussed (e.g., if an instrument demonstrates differing levels of quality depending on the country in which it is used).” (pg. 15) Additionally, we will be extracting data related to the: (i) distribution of responses in the study population and relevant subgroups; and (ii) availability of instruments for application in different settings and languages, which will also aid discussion around the accessibility of these instruments to diverse birth populations.
R26: Looking at international clinical maternity outcomes there is clear evidence of inequity, particularly in Black, Indigenous, and People of	A26: Thank you for noting these interesting points; we will keep them in mind for the

Colour (BIPOC) populations. It would be important to see how this is revealed in PROMS. Also, while the authors specifically discuss the difference between PREMs and satisfaction surveys on page 8, inequity has also been shown in satisfaction measures. Discussions of PREMs in the context of a range of populations is also vital.	discussion section of the systematic review manuscript.
R27: As the authors discuss in lines 45-54 on page 4, it is likely PREMs and PROMs will not be standard for all, not just in different settings but acknowledging the birthing population is not homogenous. The COSMIN cross-cultural validity measurement (Table 1) may go some way to address this but additional clarity would be helpful to this reader.	A27: We agree that this will be interesting to investigate. However, at this early stage of the review and not knowing exactly what we will find, we have refrained from making additional comments on the use of these measures for heterogenous populations of women.
R28: A minor comment is that there are the changes in tense in the document e.g. Page 5 Line 58 “this study will be “compared to page 6 line 32 “studies were included”	A28: Thank you, we have now gone back through to revise the tense of the document.
R29: The public and patient involvement in the project is excellent to see. I assume this would include diverse representation and cross-cultural consumer input. A comment around consumer inclusivity would be useful.	A29: We refer to the reviewer to our section on patient and public involvement (pg. 16) where we have spoken about how we have already and will continue to involve consumer representatives in the project.
Reviewer 3	
R30: I think this is a nicely written and timely piece as there is much interest in the use of peri-partum patient reported measures. The review of peripartum PREMs is novel and interesting.	A30: Thank you for taking the time to review our manuscript.
R31: My one reservation is that the authors do not describe how their work differs (or even acknowledge the publication of) several recent COSMIN publications in the postpartum space: PMID: 32442292 PMID: 34042993 PMID: 34013345 PMID: 34016441 PMID: 31429919 The authors should be strongly encouraged to explicitly state how their work is different to that published already and what gaps exist in the current literature that will be answered by their review.	A31: Initially, we would like to acknowledge the tremendous effort you and your team have committed to this topic. Our intention (which was not clearly articulated in the first iteration of the manuscript) is to identify maternity PROMs that could capture the outcomes of an entire maternal population. Thus, our interest is not in replicating the work you have already undertaken. Instead, we’re aiming to identify and appraise PROMs that capture outcomes relevant to all women across the pregnancy, childbirth and postpartum continuum (and not just women suffering postpartum depression, for example). We have revised the introduction (pg. 4-5) and eligibility criteria (pg. 7-8) to better align with this intention.

VERSION 2 – REVIEW

REVIEWER	Dawson, Pauline University of Otago, Women's & Children's Health
REVIEW RETURNED	23-Dec-2021

GENERAL COMMENTS	Thank you for the opportunity to review this resubmission. The authors have addressed my comments in their reply. I did want to make some final comments On page 5 and 6 additional wording has been added stating the PREMs and PROMs will be "relevant to all women across the pregnancy". From this I take the premise of the systematic review is identifying measures that are homogenous and can be generalised across *all* birthing people. While this is incredibly valuable, I am still not reassured that the review (as written for publication) will address heterogeneity. In A25 the authors give further excellent detail relating to different settings and subgroups, yet this detail does not seem to specifically appear in the review proposal for publication. Coming from an equity viewpoint I would like this point to be made very clear. As an example, I recently read work from a developing nation where having latrine in the birthing unit was very important to the birthing population. This is unlikely to be a PREM for say someone in Manhattan as it would be a given? In A29 the authors refer to the statement around consumer representatives on page 16 which I read in the original submission. In my initial comment I applauded this involvement however I still would like to see a statement about consumer (or researcher) diversity. For example, involvement of indigenous peoples would reassure this reviewer that while looking to find measures that can be applied to all, measures that are vitally important to select groups are not ignored in the process – as this would be inequitable. As it stands this proposal sets out to find evidence around homogenous measures and seems able to do that. I also acknowledge that at this point of the research, diversity and inclusion in the project team may not be able to be addressed. However, I think that would bring added depth to this important project and provide evidence to improve systems and services for all people utilising maternity care.

VERSION 2 – AUTHOR RESPONSE

Reviewer: 2
Dr. Pauline Dawson, University of Otago

Comments to the Author:
Thank you for the opportunity to review this resubmission. The authors have addressed my comments in their reply. I did want to make some final comments

On page 5 and 6 additional wording has been added stating the PREMs and PROMs will be "relevant to all women across the pregnancy". From this I take the premise of the systematic review is identifying measures that are homogenous and can be generalised across *all* birthing people. While

this is incredibly valuable, I am still not reassured that the review (as written for publication) will address heterogeneity. In A25 the authors give further excellent detail relating to different settings and subgroups, yet this detail does not seem to specifically appear in the review proposal for publication. Coming from an equity viewpoint I would like this point to be made very clear. As an example, I recently read work from a developing nation where having latrine in the birthing unit was very important to the birthing population. This is unlikely to be a PREM for say someone in Manhattan as it would be a given?

In A29 the authors refer to the statement around consumer representatives on page 16 which I read in the original submission. In my initial comment I applauded this involvement however I still would like to see a statement about consumer (or researcher) diversity. For example, involvement of indigenous peoples would reassure this reviewer that while looking to find measures that can be applied to all, measures that are vitally important to select groups are not ignored in the process – as this would be inequitable.
As it stands this proposal sets out to find evidence around homogenous measures and seems able to do that. I also acknowledge that at this point of the research, diversity and inclusion in the project team may not be able to be addressed. However, I think that would bring added depth to this important project and provide evidence to improve systems and services for all people utilising maternity care.

AUTHOR RESPONSE: We agree with the issues raised by reviewer on the importance of equity. However, our intent from our statement, "relevant to all women across the pregnancy" was to focus on universal aspects of maternity – not on care that would be received by subsets of women, such as those accessing abortion care, or those diagnosed with gestational diabetes who would subsequently access additional testing. We have edited this sentence to clarify this point, as it is beyond the scope of this review to assess cross-cultural relevance of measures – although we agree that this is an important area of research.